

## Toetsing

### Ontwikkelingen op het terrein van formatieve toetsing

Formatieve toetsing of formatief evalueren, beide ook wel aangeduid als *assessment for learning*, wordt als verzamelterm gehanteerd voor alle toetsingsinstrumenten, activiteiten en procedures die gebruikt kunnen worden om het leerproces van leerlingen te ondersteunen (zie het hoofdstuk over toetsing elders in deze Curriculumspiegel). Naast de enorme aandacht voor summatieve toetsing (om de opbrengsten van het leerproces in kaart te brengen) via proefwerken, schoolexamens, centrale examens en internationale toetsing zoals TIMSS en PISA is er meer en meer aandacht voor formatieve toetsing, als onderdeel van het leerproces (zie bijvoorbeeld Van den Berg & Westbroek, 2014; Castelijns, Segers, & Struyven, 2015; Noteboom, 2013;).

Tien principes voor formatieve toetsing zoals geformuleerd door de *Assessment Research Group* (ARG, 2002) benadrukken onder meer formatieve toetsing als een centraal onderdeel van klassenpraktijk, als een belangrijke professionele vaardigheid van leraren alsook het belang van motivatie voor leerlingen en het ontwikkelen van het vermogen van leerlingen om zichzelf te beoordelen (*self-assessment*) en te reflecteren. Formatief toetsen bevordert ook het leergedrag in de zin dat het leren niet alleen aan het eind voor een schriftelijke toets wordt geconcentreerd, maar juist gespreid over een langere periode. Zwakke en sterke punten worden voor de leerlingen sneller duidelijk (Saab, 2014). Daarnaast wordt formatieve toetsing vaak ook verbonden met het geven van feedback aan leerlingen (zie bv de uitgebreide review van Hattie & Timperley, 2007), met *self-assessment* en *peer learning*. Dit sluit aan bij de suggesties van Wiliam en Thompson (Nilsson & Loughran, 2013) die vijf strategieën voor effectieve formatieve toetsing hebben ontwikkeld:

- verheldering leerdoelen en delen van *criteria for success*;
- ontwikkeling van effectieve klassendiscussie, vragen en opdrachten die leeropbrengsten laten zien;
- geven van feedback om leerlingen verder te brengen;
- activeren van leerlingen als eigenaar van hun eigen leren;
- activeren van leerlingen als peer instructeur.

Ook in het eindrapport van het Platform Onderwijs 2032 (2016) wordt het belang van formatieve toetsing benadrukt. Het sluit ook aan bij het pleidooi van het Platform om niet alleen aandacht te schenken aan wat meetbaar is, maar ook aan wat 'merkbaar' is. Daarbij gaat het bijvoorbeeld om de verantwoordelijkheid die leerlingen hebben genomen, de betrokkenheid die ze hebben getoond, het zelfvertrouwen dat ze hebben ontwikkeld en ervaringen die ze hebben opgedaan (ook buiten de school). Dergelijke merkbare in plaats van meetbare kwaliteiten moeten op een andere manier worden gewaardeerd.

### Curriculaire uitdagingen op het terrein van formatieve toetsing

De aandacht voor 21e eeuwse vaardigheden en procesmatige onderwijsbenaderingen (zie 'Kennis over wetenschap' hierboven), waarbij de ontwikkeling van houdingen, denk- en werkwijzen belangrijke onderdelen zijn van de leerinhoud, vragen om formatieve toetsing. Voor de algemene kenmerken van effectieve formatieve toetsing geeft de reviewstudie van Sluijsmans, Joosten-ten Brinke en Van der Vleuten (2013) aanknopingspunten. Voor het natuurwetenschappelijke domein laten Van den Berg en Westbroek (2014) zien hoe formatieve toetsing via snelle *concept checks* met korte vragen, een grafiek of een schets vormgegeven kan worden. Voorbeelden van dit soort concept checks zijn: *Zet op volgorde van klein naar groot: cel, gen, chromosoom, atoom* (Van den Berg & Westbroek, 2014) en *Which one is the odd-one-out – bird, cat, fish, elephant? Why?* (in Black, 2013). Voor het in kaart brengen van de 'niet meetbare maar wel merkbare' resultaten van onderwijs, zowel summatief als formatief, zullen mogelijkheden verzameld, ontwikkeld en beschikbaar gemaakt moeten worden. Dit is een nog grotendeels onontgonnen terrein, al zijn er enkele aanzetten. 'Merkbaar' zou kunnen worden uitgewerkt in hoe je dan iets kunt merken in termen van observeerbaar gedrag of een prestatie. Uit een eerste onderzoek dat Cito in 2008 heeft gedaan in een project met Platform Midden Nederland op een aantal vmbo scholen blijkt dat bij een vaardigheid als samenwerken leraren en leerlingen na enige oefening tot een behoorlijke overeenstemming kwamen over hoe gevorderd iemand was in samenwerken. Deze

Bron: *Natuurwetenschappelijke vakken, Vakspecifieke trendanalyse 2017*; te raadplegen via <http://natuurenteknik.slo.nl>

vaardigheid was uitgewerkt met concrete gedragsindicatoren en een vierpuntschaal. Een en ander was gebaseerd op de Protocol Portfolio Scoring van Straetmans (zie <http://www.toetswijzer.nl/html/literatuur/pps.pdf>).

Daarnaast is het belangrijk om te investeren in de toerusting van leraren. Die toerusting betreft het kunnen inzetten van formatieve toetsingsinstrumenten en -procedures en het kunnen benutten van toetsgegevens ter ondersteuning en inrichting van het verdere leerproces van leerlingen. Tevens is het belangrijk oog te hebben voor de inhoudelijke afstemming tussen toetsing en curriculum en voor de implicaties van formatieve toetsing voor curriculumbeslissingen op klas- en schoolniveau.

### Ontwikkelingen op het terrein van toetsing van vaardigheden

Als vaardigheden belangrijker worden, neemt ook het belang van de toetsing ervan toe. Die trend is ook zichtbaar in internationale onderzoeken als PISA en TIMSS, waarin meer aandacht voor vaardigheden is. Dat lijkt een kentering ten opzichte van de kleinere plaats die toetsing van vaardigheden in de schoolexamens voor de natuurwetenschappelijke vakken kreeg in 1997. Want hoewel vaardigheden nog steeds onderdeel vormen van de examenprogramma's, zijn toen de voorschriften vervallen met betrekking tot het aandeel van vaardigheidstoetsen in de schoolexamens. Dat leidde op scholen tot een afname van het aantal practica in de lessen (Van Goor & Den Hertog, 2001). In een enquête in het kader van de kwaliteit van schoolexamens in 2013 meldde echter ongeveer de helft van de biologiedocenten dat ze veel vaardigheden regelmatig toetsen, met uitschieters naar boven (informatie- en onderzoeksvaardigheden, vaardigheden met betrekking tot het wisselen van organisatieniveau) en naar beneden (ontwerpvaardigheden en vaardigheden rond studie en beroep). In Engeland is er sprake van toenemende kritiek op het huidige systeem waarin veel praktische vaardigheden in streng gereguleerde *School Based Assessments* getoetst worden. De trend is daar een van afnemende regeldruk (zie onder andere Evans, 2013; Watts, 2013), opdat er meer ruimte komt om van practica te leren. In het kader van de vernieuwingen van de examenprogramma's voor de tweede fase wordt in het DUDOC-programma onder andere onderzoek gedaan naar het (zelf) beoordelen van vaardigheden voor het doen van onderzoek (Van der Jagt, Van Rens, Schalk, Pilot & Beishuizen, 2013).

De grotere aandacht voor het toetsen van vaardigheden beperkt zich echter niet tot practicumvaardigheden. Ook hogere denkvaardigheden zoals redeneren worden expliciet benoemd in de examenprogramma's. Deze vaardigheden worden deels getoetst in de centrale examens. Cito doet op dit moment onderzoek naar een classificatie van de examenopgaven volgens enkele veel gehanteerde gedragstaxonomieën. Uit een eerste analyse blijkt dat examens ook opgaven bevatten die een beroep doen op hogere ordevaardigheden zoals analyseren, evalueren en creëren.

Er lijkt in het onderwijs ook meer oog voor een geïntegreerde aanpak van leren en toetsing. In de Verenigde Staten leggen de nieuwe *Next Generation Science Standards* (National Research Council, 2012) de nadruk op integratie van kernbegrippen ('*Core ideas*'), denkwijzen ('*Crosscutting Concepts*') en werkwijzen ('*Practices*') van de natuurwetenschappen. Het leerplanvoorstel N&T onderbouw vo (Ottevanger e.a., 2014) bouwt voort op dit gedachtegoed. Het '*driedimensionale leren*' dat hierbij hoort, wordt ook voor de toetsing bepleit (National Research Council, 2014). Toetsing zou zich op meer aspecten moeten richten dan kennis en praktische vaardigheden. Om te bepalen in hoeverre leerlingen zich de kernbegrippen, denk- en werkwijzen van de natuurwetenschappen eigen hebben gemaakt, zou een mix van verschillende toetsinstrumenten moeten worden ingezet. Ook in Nederland krijgt breder evalueren steeds meer aandacht (Castelijns et al., 2011; Joosten-ten Brinke, 2011; Sluijsmans, 2008). In het beroepsonderwijs heeft het beoordelen van competenties op basis van portfolio's vaste grond onder de voeten gekregen (Oosterheert, Van Eldik, & Kral, 2007). In havo en vwo staat dat nog in de kinderschoenen. Een voorloper op dat terrein zijn de technasia waar naast gebruik van portfolio's (Ketelaar, 2013) ook op andere manieren geïntegreerde toetsing van opgedane kennis, vaardigheden en professionele ontwikkeling plaatsvindt (Schalk & Bruning, 2014). Daarbij is aandacht voor evenwicht tussen het beoordelen van proces en product en tussen beoordelen van het geheel en de delen.

### Curriculaire uitdagingen op het terrein van toetsing van vaardigheden

Meer aandacht voor vaardigheden betekent meer ruimte ervoor in de lessen, ze explicieter onderwijzen en ze explicieter beoordelen. Om het belang ervan te onderstrepen zou een (nog) explicietere vermelding van vaardigheden in doelen, examenvoorschriften en PTA's wenselijk zijn. Maar leraren vinden het toetsen van vaardigheden vaak lastig. Hoewel er op veel scholen een praktijk gegroeid is van bijvoorbeeld het beoordelen van praktische opdrachten en profielwerkstukken, blijven leraren onderzoeksvaardigheden noemen als meest moeilijk te beoordelen categorie vaardigheden waarbij ze graag ondersteuning zouden willen. SLO werkt aan die ondersteuning door gebenchmarkte beoordelingsinstrumenten op te stellen, zodat docenten kunnen zien wat hun collega's op dit punt doen en belangrijk vinden. Zij kunnen dan een gefundeerde keuze maken voor welke criteria zij zelf gebruiken in verschillende fasen van het leerproces. Andere voor ondersteuning van docenten hoog scorende vaardigheden zijn redeneren, reflecteren en waarderen. Ook daarvoor lijkt een soortgelijke aanpak gewenst. SLO werkt aan een 'SPA+', een systematische probleemaanpak voor redeneren, die ook als basis kan dienen voor de beoordeling ervan. Met betrekking tot de denk- en werkwijzen uit het leerplanvoorstel N&T voor de onderbouw vo zou ingezet moeten worden op het ontwikkelen van (voorbeelden van) de genoemde mix van toetsvormen.